

CORPORATE CONFIDENTIAL

SPEECHGEAR, INC.

Progress Report

Period Covered by the Report

• 10 September 2001 to 9 November 2001

Date of Report

• 10 November 2001

Project

• *Compadre: A Device Independent Voice-to-Voice Language Translator Software Solution*

• SBIR Phase I Topic N01-044

• Contract Number N00014-01-M-0225

Item Number

0001AC: Progress Report

Security Classification

Unclassified

Sponsor

Dr. Joel Davis
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660



Principle Investigator:

Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021



SPEECH GEAR

Program Partners



VISUAL
GOLD



IRISUSA

Aramedia

Software Utilization Technology Group

ARAMEDIA



L&H/Aptek



CiMOS

A. Project Summary

Technical Abstract:

Mission Statement

To develop and deploy language translation software that is device independent, supports bi-directional translation of multiple languages, produces text transcriptions of spoken conversations and supports translation of text extracted from digital images. This software shall run in both a reduced functionality standalone mode, and by wirelessly connecting to remote servers, a full-function mode. This software shall run on multiple pocketable platforms resulting in a mobile system that is low in cost, easy to use, robust in operation and comfortable to carry, and/or wear.

The object of this Phase I research effort is to investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement. Specifically, the team will investigate design options for the mobile translator system, identify potential applications, and select the best option(s) to pursue in making the design a reality. Four technical areas will be investigated: potential pocketable computing platforms, the operator interface, optical character recognition software and the language translation software. The commercial feasibility of this design will also be investigated. This includes identifying potential applications, languages to be supported, cost, and user requirements such as interface modes and response times. By combining both the commercial and technical elements, a complete definition of successful software and system solutions for pocketable language translation devices will be achieved.

Prototype systems showing device independence will be developed and demonstrated and a final report written documenting the Phase I results and recommendations for follow-on research and development in Phase II. Options are included for incorporating additional language pairs into the system and application specific terminology.

Anticipated Benefits/Potential Commercial Applications of the Research or Development:

Applications include all individuals who require multi-lingual capabilities. The mobile translator will benefit a wide range of individuals including military personnel, airport employees, border patrol and customs agents, police, fire fighters, retail clerks, bank tellers, delivery personnel, phone operators, tourists and any industry that sells, develops or manufactures products to/in global markets or employs individuals that do not speak the native language.

B.3 Camera-Based Mode

The primary means to input text into the SmartPhone for this mode of usage will be a digital camera. A patent application for this capability has been submitted. The design of the prototype system is shown in Figure 5. Two different cameras are being used: a compact camera from HP and a high resolution camera from Minolta. The Minolta *Dimage 7* is being used to perform the initial testing for *Compadre*. Once this camera has been successfully integrated and tested, then SpeechGear will proceed to integrate and test lower resolution cameras such as the HP camera that is shown.

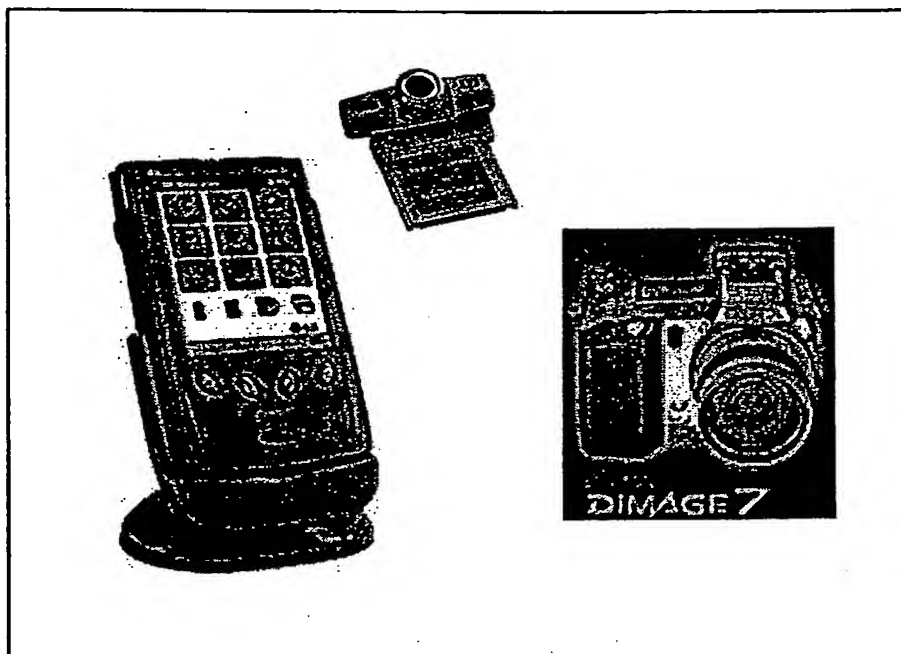


Figure 5: Examples of Camera-based Systems

Note that *Compadre*'s software is designed to be device independent, thus, these are just two of many hardware configurations that could be used for this usage mode. One interesting alternative device is Samsung's conceptual product of including a camera with a cellular phone. This product is shown in Figure 6.

The digital camera will be used to capture an image of the foreign language. Such a picture is shown in Figure 7. Once the desired image is obtained, the SmartPhone will wirelessly connect to a remote server where the image will be processed and the resulting translation sent back to the user. An example of the translated text in the "one-click" GUI is shown in Figure 7. For most applications, this connection will be made using cellular telephones. Because of the limited bandwidth of such a connection, it is important to reduce the overall size of the transmission. Thus, SpeechGear evaluated

different image compression algorithms and selected the *Imagist* product from Visual Gold. *Imagist* will be embedded directly into SpeechGear's software, and thus will be transparent to the end user.

The GUI being developed for *Compadre* is shown in Figure 7. After capturing the image(s), the user will simply select "Translate" and the wireless connection will automatically be established. Note that multiple images can be sent simultaneously using a single click. This is similar to the "Add to Basket" interfaces that are being used at web-based shopping sites. In this approach, selected items are loaded into a virtual basket or cart, and once you are done shopping you select "Check Out" to purchase all of the items simultaneously. For *Compadre*, multiple images can be selected and entered into the queue, and when the user is ready to connect to the remote server, then simply selecting the "Translate" button will connect the

SmartPhone to the remote server, which in turn will process the images and return the resulting translation. The images will be transmitted back to the user using an HTML format. The users can then scroll through these images and save or delete them as is desired. Please note that the actual buttons will be Icons versus text, and thus the look and feel of the resulting GUI will be a substantial improvement over what is shown in the figures.

One item of note is that *Compadre*'s Hybrid Translator can be configured to handle different types of input using a variety of methods. For voice-based input, the context in which words are used is readily available. This often is not the case with the camera-based mode. For example, the words "Post Office" without context could be interpreted as a "Pole that is stuck in the ground" and "A place where people work." Thus, SpeechGear is configuring the translator to be dominated by a Translation Memory (TM) mode versus Machine Translation (MT). In TM, the translator uses a known set of previously translated phrases to achieve accurate outputs. Such an approach is used very often if for example an operator's manual has been previously translated, but has now been updated and thus needs to be translated once again. In the case of the camera-based system, the TM approach will be used to enter signs and information, such as the Post Office example that was stated above. Thus, SpeechGear is in the process of building the TM database to include signage typically seen on signs.

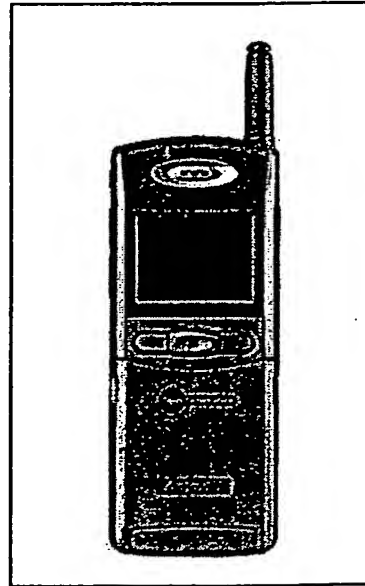


Figure 6: Samsung's Proposed Combined Camera and Digital Cellular Phone

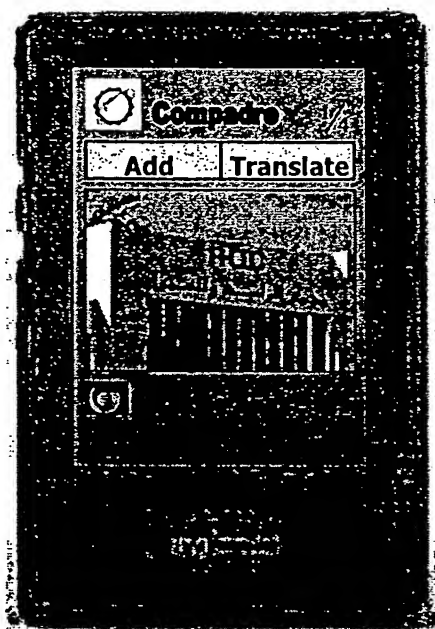


Figure 7: Preliminary Graphical User Interface to Submit Images for Translation

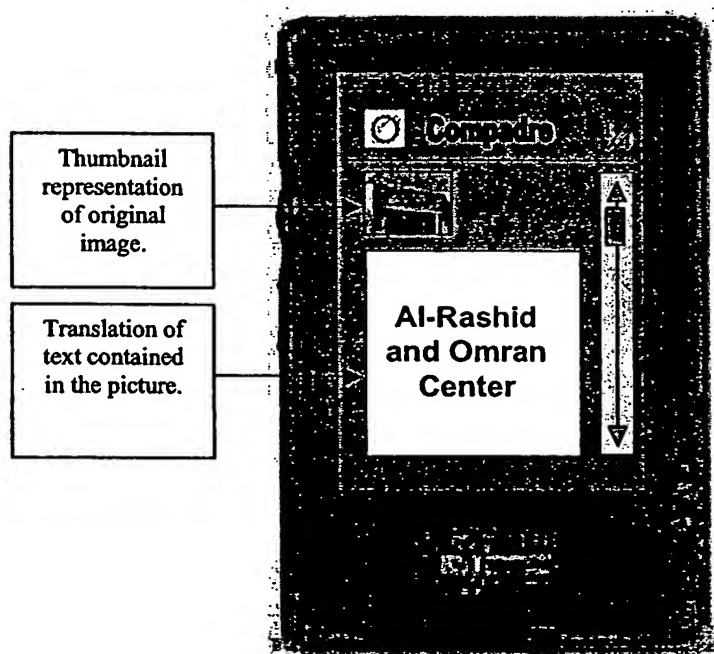


Figure 8: Preliminary Graphical User Interface for Viewing Results of Translation